

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.383

Подвальников
Дмитрий Сергеевич

Распознавание рукописных математических выражений с изображения

АВТОРЕФЕРАТ

на соискание академической степени
магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель
Анисимов В.Я.
кандидат физико-математических
наук, доцент

Минск 2019

КРАТКОЕ ВВЕДЕНИЕ

С течением времени становится всё больше различных научных исследований, в которых используется большое количество математических формул, так как через неё можно легко выразить сложные вещи. В настоящее время уже много готовых математических выражений, которые требуется вставить в документ, содержится в технических справочниках и других документах или научных работах.

Вставка математических выражений в документы гораздо сложнее, чем можно предположить. Когда человеку необходимо вставить математическое выражение или оцифровать рукописную формулу в какой-либо документ, он должен быть знаком с набором команд LaTeX, интерфейсом Microsoft Word или другими инструментами, позволяющими осуществлять ввод математических выражений. Независимо от того необходимо набрать формулу с нуля или же немного изменить её, человеку придётся набирать формулу всю целиком, используя один из инструментов, предназначенных для этих целей. Набор больших математических выражений очень объёмная работа, которая требует значительные затраты времени. Таким образом, автоматизация процесса ручного набора математических выражений, путём распознавания математической формулы с изображения и конвертации её в удобно читаемый формат, позволила бы экономить время на создание и поддержку научных документов, содержащих большое количество математических выражений, а также экономить на тренингах по обучению сотрудников различным инструментам, позволяющим осуществлять работу с математическими формулами.

Автоматическое распознавание рукописного или машинописного математического выражения было важной областью исследований на протяжении многих лет. Как правило, распознавание математического выражения включает в себя три основных этапа: сегментация изображения, распознавание и структурный анализ. Сегментация изображения используется для получения отдельных математических символов и обозначений. С помощью распознавания можно классифицировать полученные отдельные символы в результате сегментации. Структурный анализ отвечает за исследование пространственных связей между символами для определения наиболее вероятной взаимосвязи отдельных символов (отношения и местоположения символов) из которого в конечном итоге будет воспроизведена формула.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является исследование и реализация алгоритмов в области сегментации изображений, распознавании символов с изображения и подходов в структурном анализе.

Объектом исследования являются изображения, содержащие математические символы и формулы.

Предметом исследования являются алгоритмы сегментации, структурного анализа и машинного обучения.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично при изучении и анализе научных работ.

Опубликованность результатов диссертации

Не публиковалась.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, списка публикаций автора. В первой главе произведен обзор теоретической части. Вторая глава посвящена используемым технологиям, в частности дается всех используемых технологий для достижения результата. Третья глава содержит непосредственно модулей системы, позволяющей распознавать математические выражения. В заключении подводится краткий итог всех примененных оптимизаций.

Общий объем работы составляет 70 с., 22 рис., 2 табл., 3 формулы, 30 источников.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** рассказывается о важности роли математических формул в нынешнее время и о сложностях, с которыми можно столкнуться при наборе этих формул в различных редакторах. Кратко описывается, как может помочь автоматическое распознавание математических формул и даётся краткое описание диссертации.

В **первой главе** даётся описание теоретических сведений: существующих подходов в сегментации изображения, направления, сферы использования и алгоритмы машинного обучения, методы структурного анализа математического выражения. Кратко описывается принцип работы нейронных сетей.

Вторая глава посвящена используемые технологии. Рассказывается про язык программирования Python, на котором были разработаны модули системы распознавания и интерактивную вычислительную среду Jupyter Notebook. Дается обзор таких библиотек как NumPy, Matplotlib, Pandas и Keras для языка программирования Python. Также приводятся преимущества программной библиотеки с открытым исходным кодом OpenCV, которая содержит алгоритмы обработки изображений, компьютерного зрения и машинного обучения.

Третья глава рассказывает о цели всей работы – этапах исследования и реализации программного модулей системы для распознавания математических выражений. На практике рассматриваются конкретные подходы в сегментации и структурном анализе. Проводится анализ и сравнения для выбора лучших параметров нейронной сети для распознавания математических символов.

Вначале описывается модуль сегментации изображения, содержащего математическую формулу. Рассказывается о бинарном представлении изображения, его сегментации и с какими проблемами можно столкнуться при сегментации отдельных объектов на изображении.

Затем проводится обзор обучающего набора данных математических символов для распознавания отдельных символов, полученных в результате сегментации изображения. Проводится сравнение обычного градиентного спуска, пакетного градиентного спуска и стохастического градиентного спуска. Обучение нескольких моделей нейронных сетей с различными

архитектурами (различное количество скрытых слоев, различные функции активации)

В последней части третьей главы рассказывается про выбор структур данных для структурного анализа математического выражения. Рассматривается подход для восстановления положения символов в изначальный вид, которых был на изображении.

В главе **Заключение** подытоживаются полученные результаты по проделанной работе и полученному результату. Рассказывается, каким образом можно попытаться улучшить текущую реализацию.

ЗАКЛЮЧЕНИЕ

В диссертации был рассмотрен вопрос распознавания рукописного математического выражения с изображения и восстановления его структуры. В рамках диссертации была разработана программа, состоящая из трёх модулей, которые взаимодействуя между собой, предоставляют данные возможности.

Были исследованы различные техники сегментирования изображения, алгоритмы машинного обучения для классификации изображений и подходы по восстановлению структуры математического выражения. Для сегментации изображений были произведены сравнения работы алгоритмов при различных условиях фотосъемки и выбран наилучший подход в совокупности по скорости и качеству работы. При выборе алгоритма распознавания во внимание принимались точность предсказания и скорость обучения нейронных сетей. Главными критериями при реализации модуля структурного анализа были использование структуры данных, из которой было бы удобнее всего восстановить структуру математического выражения и соответствие структуры полученной в результат работы исходной структуре выражения. В целом результаты работы программы стоит признать удовлетворительными.

В результате цель диссертации достигнута. Был написан программный код. В дальнейшем планируется развивать существующее приложение путём расширения существующего набора данных как новыми математическими символами для распознавания более широкого спектра математических символов, так и увеличением числа уже существующих для более точных предсказаний. Также путём дальнейшего изучения, проработки и применения, более сложных и требующих большей вычислительной мощности алгоритмов сегментации изображения и структурного анализа, планируется добиться лучшего качества конечного результата при различных побочных эффектах, оказываемых на изображение.

В части сегментации текущие алгоритмы могут давать неудовлетворительный результат при плохом освещении или плохом качестве изображения. Поэтому необходимо проверить технику Transfer learning для обнаружения отдельных символов математического выражения используя при этом такие архитектуры нейронных сетей как: YOLO, R-CNN, Fast R-CNN, R-FCN.

Для распознавания отдельных математических символов планируется попробовать более сложные архитектуры нейронных сетей, для которых

необходимы большие вычислительные мощности. В частности планируется использовать свёрточные нейронные сети.

Для структурного анализа также планируется посмотреть другие подходы, связанные с машинным и глубоким обучением.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1-А. Евжик Д.А., Подвальников Д.С., Тишковский М.А. Сравнение вариаций градиентного спуска на примере задачи распознавания символов / Евжик Д.А., Подвальников Д.С., Тишковский М.А // 55-я юбилейная научная конференция аспирантов, магистрантов и студентов. – 2019. – № 4(20). – с. 195-197.

2-А. Тишковский М.А., Лимонтов А.С., Подвальников Д.С. Оценка алгоритмов машинного обучения. Дилемма смещения-дисперсии / Тишковский М.А., Лимонтов А.С., Подвальников Д.С. // 55-я юбилейная научная конференция аспирантов, магистрантов и студентов. – 2019. – № 4(20). - с. 221-223.